

Die internationale Einbettung des Sozio-oekonomischen Panels (SOEP) im Rahmen des Cross-National Equivalent File (CNEF)*

von Joachim R. Frick**, Stephen P. Jenkins, Dean R. Lillard,
Oliver Lipps und Mark Wooden

1 Hintergrund

In den vergangenen 30 Jahren haben mehrere Industrienationen zunehmend in Entwicklung, Aufbau und Durchführung von sozialwissenschaftlichen Haushalts- und Personenbefragungen investiert, wobei Längsschnitt- und Panelstudien eine besondere Rolle zukommt. In vielen Fällen von der Wissenschaft vorangetrieben, unterstützen diese Studien die Erfassung, Beschreibung und Analyse sozio-ökonomischer Merkmale der Bevölkerung eines Landes und die Beobachtung von Lebensverläufen. Derartige „general purpose social science surveys“ ermöglichen außerdem die Erfassung von Veränderungen des Verhaltens im Zusammenhang mit sozialpolitischen Neuerungen und Veränderungen. Es können darüber hinaus Szenarien zur Abbildung möglicher Verhaltensvariationen unter alternativen sozialpolitischen Maßnahmen entwickelt werden. Diese Daten haben die sozial-, wirtschafts- und verhaltenswissenschaftliche Forschung nicht nur auf nationaler Ebene gestärkt, sondern auch länderübergreifende Studien vorangetrieben.

Um nationale Surveydaten für international vergleichende Analysen nutzen zu können, muss untersucht werden, inwieweit die Daten überhaupt vergleichbar sind bzw. vergleichbar gemacht werden können. Dies erfordert gegebenenfalls umfangreiche methodische Maßnahmen, meist zur Anpassung bereits vorhandener Daten im Rahmen der so genannten Ex-post-Harmonisierung. Da die Befragungen in den einzelnen Ländern mit Hinblick auf nationale Politik und nationale Forschungsziele durchgeführt wurden, sind sie im Allgemeinen nicht von vornherein dafür entworfen worden, für international vergleichbare Studien geeignete Daten zu generieren. Ausnahmen bilden hier die Mikrodaten des von Eurostat, dem Statistischen Amt der EU, für die EU-15-Staaten koordinierten Europä-

* Das Cross-National Equivalent File (CNEF) wurde 1991 von Richard V. Burkhauser (damals Maxwell School der Syracuse University, New York/USA) und der SOEP-Gruppe am Deutschen Institut für Wirtschaftsforschung (DIW Berlin) ins Leben gerufen, damals weitgehend finanziert vom National Institute of Aging (USA). Inzwischen ist es ein gemeinschaftliches Projekt von sechs Institutionen: Cornell University (wohin R. V. Burkhauser zwischenzeitlich wechselte), SOEP am DIW Berlin, Statistics Canada in Ottawa, Institute for Social and Economic Research (ISER) an der Universität von Essex, Melbourne Institute of Applied Economic and Social Research an der Universität von Melbourne und der Universität Lausanne (bis 2007 Universität Neuchâtel). Die Autoren danken insbesondere Richard V. Burkhauser, Gaëtan Garneau, Robert Schoeni und Gert G. Wagner für hilfreiche Kommentare zu früheren Entwürfen dieses Artikels.

** DIW Berlin, TU Berlin, IZA Bonn, E-Mail: jfrick@diw.de

ischen Haushaltspanels ECHP (European Community Household Panel) und die Kohortenstudie SHARE (Survey of Health, Ageing and Retirement in Europe). Im Rahmen der beim ECHP praktizierten Input-Harmonisierung wurde bereits bei der Erhebung explizit die spätere Verwendung der Daten für international vergleichende Analysen berücksichtigt, indem abgestimmte Erhebungsinstrumente mit möglichst ähnlicher Frageformulierung verwendet wurden. Das ECHP war jedoch nur zum Teil erfolgreich und wurde nach acht Wellen im Jahre 2001 aufgegeben. SHARE ist erfolgreicher, aber aufgrund seiner Fokussierung auf ältere Kohorten für die Analyse breiterer Bevölkerungsschichten oder gar der Gesamtbevölkerung ungeeignet.¹

Die meisten Daten müssen jedoch nachträglich harmonisiert werden. Hierbei ist für die Definition von Variablen zur Erfassung äquivalenter Konstrukte und Verhaltensweisen in der international komparativen Forschung erheblicher Aufwand notwendig. Im Zusammenhang mit einfachen Konzepten wie Alter und Geschlecht ist diese Aufgabe vergleichsweise unkompliziert. Erheblich komplexer wird es, wenn es sich um Konzepte handelt, die nur im Kontext von landesspezifischen Institutionen und kulturellen Faktoren verstanden werden können. International vergleichbare Indikatoren zur Messung von Faktoren wie ökonomische Wohlfahrt, Bildung, Beschäftigung und Gesundheit können oft nur nachträglich gebildet werden, da die in den jeweiligen Ländern erfassten Daten einen landesspezifischen institutionellen Kontext haben oder sich auf kulturelle Faktoren begründen. Für die vergleichend arbeitenden Forscher ist es daher wichtig, sich mit den Institutionen, Gesetzen und kulturellen Gegebenheiten der verschiedenen Länder vertraut zu machen.

Eine der ersten Studien, die international vergleichbare Daten bereitstellte, war die Luxembourg Income Study (LIS). Seit 1983 harmonisiert LIS Mikrodaten aus Erhebungen aus mehr als 30 Ländern (vgl. www.lisproject.org und Smeeding, Jesuit und Alkemade 2002), darunter auch für Deutschland auf Basis der Daten des SOEP. Da die Hauptlast der Harmonisierung („Lissification“) bei LIS liegt, wird der Aufwand für den einzelnen Forscher erheblich reduziert.

Obgleich der Umfang der LIS-Aktivitäten beeindruckend ist, können damit aber nicht alle Bedürfnisse der internationalen Forschergemeinschaft erfüllt werden. So erhalten externe Wissenschaftler über das Datenfernverarbeitungssystem LISSY nur indirekten Zugang zu den der LIS-Datenbank zugrunde liegenden Mikrodaten, da es sich in einigen Fällen um nicht direkt weitergabefähige Datenbestände der nationalen Statistischen Ämter handelt. Obwohl eine umfangreiche Dokumentation der LIS-Daten vorliegt, hat dieser Ansatz zur Folge, dass Forscher sich mit der vorgegebenen Harmonisierung „zufrieden geben“ müssen und deren Einfluss auf substantielle Analysen nicht prüfen können. Zudem liefert LIS nur Querschnitt-Daten, so dass an Panel- oder Längsschnittanalysen interessierte Forscher eher wenig von den LIS-Daten profitieren können.

Im Folgenden wird ein Projekt vorgestellt, das zwar auf der LIS-Idee aufbaut, deren Beschränkungen jedoch weitgehend überwinden kann. Bei diesem Projekt, dem Cross-National Equivalent File (CNEF), handelt es sich um ein gemeinsames Unterfangen von Wissenschaftlern und Vertretern von Institutionen aus sechs verschiedenen Ländern (Stand: Sommer 2008), die jeweils nationale Panel-Erhebungen durchführen. Hierzu gehören die

¹ Burkhauser und Lillard (2005) beschreiben detailliert die Erfolge und Misserfolge bei der Erstellung von sowohl ex-ante als auch ex-post harmonisierten Datensätzen für die international vergleichende Forschung im Allgemeinen; Lillard und Burkhauser (2006) beziehen sich insbesondere auf SHARE.

US-amerikanische *Panel Study of Income Dynamics* (PSID), das *Sozio-ökonomische Panel* (SOEP) aus Deutschland, der *British Household Panel Survey* (BHPS) aus Großbritannien, der *Survey of Labour and Income Dynamics* (SLID) aus Kanada, der *Household Income and Labour Dynamics Survey* (HILDA) aus Australien sowie das *Schweizer Haushalt-Panel* (SHP). Die Koordination der CNEF-Aktivitäten liegt in den Händen von Richard V. Burkhauser und Dean R. Lillard an der Cornell Universität in Ithaca, NY (USA).

Im CNEF werden ausgewählte Variablen harmonisiert, die in mindestens zwei der verfügbaren nationalen Datenbestände enthalten sind, die ihrerseits zu den weltweit erfolgreichsten, laufenden sozialwissenschaftlichen Panelstudien gehören. Externe Wissenschaftler haben im Prinzip Zugang sowohl zu den harmonisierten wie auch den Originaldaten, die sie auf der Individualebene mit den harmonisierten Daten zusammenführen können, sowie einer umfangreichen Dokumentation der dem Harmonisierungsprozess zugrunde liegenden Annahmen und weiteren Hinweisen zur Vergleichbarkeit der Variablen.

Das CNEF unterscheidet sich von anderen Standardisierungs-Projekten nicht nur, weil es Daten von laufenden Erhebungen mit einbezieht, sondern auch dadurch, dass die wissenschaftlich arbeitenden Nutzer direkt Einfluss auf die Auswahl neuer CNEF-Variablen sowie deren Harmonisierung ausüben können. Solche Variablen entstehen häufig in Forschungsprojekten, in denen Wissenschaftler international vergleichbar arbeiten. Das dabei erworbene Expertenwissen auf dem jeweiligen Forschungsgebiet beinhaltet auch Kenntnisse von landesspezifischen Institutionen. Dies hat zur Folge, dass die im CNEF enthaltenen harmonisierten Daten das Wissen vieler Wissenschaftler aus verschiedenen Disziplinen vereinen. Da die Forscher ihre harmonisierten Variablen ständig verfeinern sowie neue Variablen hinzufügen, entwickelt sich der Datenbestand des CNEF kontinuierlich weiter.

Einer der großen Vorteile des CNEF liegt im Panel-Charakter der Daten, die es erlauben, eine Person über viele Jahre hinweg zu verfolgen. Dies ermöglicht mithilfe geeigneter statistischer Methoden die Analyse von Stabilität und Wandel individueller Verhaltensweisen, was mit wiederholten Querschnittbefragungen nicht beobachtbar ist. Die im CNEF vereinten Längsschnittstudien ermöglichen somit die Untersuchung des Einflusses institutioneller Veränderungen in interregionaler und intertemporaler Hinsicht. Dies führt zu einem besseren Verständnis des menschlichen Verhaltens. Insbesondere bei lang laufenden Studien wie der PSID und dem SOEP ermöglichen die nationalen Befragungsdaten die Beobachtung von Familien über mehrere Generationen hinweg. Insofern sind die Daten des CNEF zur Untersuchung von z. B. sozioökonomischem Status und dessen Übertragung von Generation zu Generation, in internationaler wie in interdisziplinärer Perspektive, bestens geeignet (siehe hierzu z. B. Butz und Torrey 2006).

2 Die Entwicklung des CNEF

Die Geschichte des CNEF begann 1991 mit einer harmonisierten Datenbasis für lediglich zwei Länder, den USA und Deutschland, gefördert durch das National Institute of Aging².

² Die Principal Investigators dieser Initiative waren Richard V. Burkhauser, zu dieser Zeit an der Maxwell School der Syracuse University, NY (USA) und Gert G. Wagner, Direktor des SOEP am DIW Berlin. Besonderer Dank geht an Richard Hauser, zu dieser Zeit an der Universität Frankfurt/Main, für seine außerordentlich wichtige Unterstützung in der Initialphase dieses Projektes.

Im Jahr 1999 kamen die Daten des BHPS aus Großbritannien und des SLID aus Kanada hinzu. 2007 folgte der HILDA Survey aus Australien und es begannen die Harmonisierungsarbeiten für die Daten der Schweizer Haushaltspanel-Studie SHP, sodass das CNEF heute harmonisierte Datensätze von sechs Ländern beinhaltet.

Der ursprüngliche Fokus des CNEF lag auf dem Aufbau einer harmonisierten Datenbasis für die vergleichende Analyse von einkommensbasierter Ungleichheit sowie Einkommensmobilität in den USA und Deutschland mithilfe der repräsentativen Mikrodaten von PSID und SOEP. Viele der frühen vergleichenden Arbeiten zu den Themen ökonomische Wohlfahrt und Einkommensmobilität in den USA und Deutschland in den 1980er und 1990er Jahren basierten auf diesen harmonisierten Daten (vgl. z. B. Burkhauser, Frick und Schwarze 1997, Burkhauser, Crews-Cutts und Lillard 1999). Im Rahmen des weiteren, forschungsgetriebenen Ausbaus des CNEF wurden zusätzliche inhaltliche Bereiche berücksichtigt, wobei insbesondere die in den nationalen Befragungen zunehmend umfassendere Abbildung des Bereichs Gesundheit auch für das CNEF eine bedeutende Rolle spielt.

Der anhaltend hohe Forschungsbedarf zum Thema *ökonomische Ungleichheit* im Rahmen wohlfahrtsstaatlicher Vergleiche und der Suche nach „best practise“ führte fast zwangsläufig zu einer Erweiterung des CNEF um weitere nationale Datensätze. Für den Erfolg des CNEF-Ausbaus war dabei nicht nur wichtig, dass in den zwischenzeitlich neu aufgelegten Panelstudien BHPS und SLID das Einkommen von Individuen und Haushalten einen Befragungsschwerpunkt darstellte, sondern dass beide Befragungen in ihrem Aufbau von den Erfahrungen des PSID und des SOEP beeinflusst waren, was der Vergleichbarkeit der Mikrodaten sehr dienlich war. Auf Basis dieser erweiterten Datenbasis entstanden eine Reihe von vergleichenden Studien zur Entwicklung von Ungleichheit, Löhnen und (Einkommens-)Mobilität in den USA, Kanada, Großbritannien und Deutschland (vgl. z. B. Jenkins und Schluter, 2003; Jenkins, Schluter und Wagner 2003; Burkhauser et al. 2005). Im Laufe der Zeit wurden zusätzliche Variablen harmonisiert. Eine umfangreiche inhaltliche Erweiterung der CNEF-Variablen fand im Jahr 2003 statt, als Variablen zum Thema Gesundheit, die in mindestens zwei der damals teilnehmenden Längsschnittstudien vorlagen, harmonisiert wurden (vgl. Lillard und Burkhauser 2005). Daten aus der HILDA Befragung kamen im Jahr 2007 hinzu und unterstützen international vergleichende Panel-Analysen zu den Themen Beschäftigungschancen und Erwerbseinkommen von Arbeitnehmern mit Behinderung (vgl. Burkhauser, Schmeiser und Schroeder 2007). Zum Zeitpunkt der Drucklegung dieses Artikels werden nach entsprechenden Vorarbeiten auch Daten des SHP aus der Schweiz einbezogen. Die nächste Version des CNEF wird somit voraussichtlich die folgenden Daten enthalten: PSID (Wellen 1980–2005), SOEP (1984–2006), BHPS (1991–2005), SLID (1992–2005), SHP (1999–2005), sowie HILDA (2001–2005). Die in den jeweiligen Studien analysierbare Zahl der Beobachtungen auf der Ebene von individuell befragten Erwachsenen und deren minderjährigen Kindern bis 2005 findet sich in Anhang 2. In der Summe aller sechs Befragungen ergeben sich dabei weit über 2,5 Millionen Personen-Jahr-Beobachtungen.

3 Aufbau und Inhalt des CNEF

Das Ziel des CNEF ist es, unabhängig von den Erfahrungen der jeweiligen Forscher im Umgang mit Längsschnittdaten, international vergleichende Analysen mit wirtschafts- und

sozialwissenschaftlichen Daten zu unterstützen. Zur Erreichung dieses Ziels, das sowohl Forschung als auch „capacity building“ beinhaltet, stellt das CNEF Mikrodaten mit möglichst benutzerfreundlich definierten Variablen aus verschiedenen Surveys zusammen. Für jeden nationalen CNEF-Datensatz existiert eine separate Datenbank, in der die entsprechenden Variablen wellen- bzw. kalenderjahrspezifisch abgelegt werden. Diese können dann entweder im Sinne von *Stand-alone*-Daten oder, wie in den meisten Fällen, in Kombination mit anderen interessierenden Informationen, analysiert werden. So werden z. B. landes- und zeitspezifische Daten zu politischen und ökonomischen Rahmenbedingungen wie Arbeitslosenquoten mit den CNEF-Mikrodaten zusammengeführt. Da die Identifikator-Variablen des CNEF für alle Beobachtungseinheiten jenen der Originaldatensätze entsprechen, können durch entsprechende Verknüpfungen weitere Mikrodaten aus einer oder mehrerer der ursprünglichen Datenbanken in die CNEF-Datenbank überführt werden, wobei diese als eine Art zentrale Navigationsdatei oder *Masterfile* fungiert.

Der Aufbau des CNEF erleichtert weniger erfahrenen Forschern die Arbeit, da die Variablen in jedem Datensatz mit identischen Variablennamen und – soweit möglich – Wertbezeichnungen versehen sind. Die Namen der Variablen spiegeln ihren Inhalt wider: Der erste Buchstabe bezeichnet die Kategorie der Variablen: demographic (D), employment (E), household composition (H), income (I), weighting (W), sample identifiers (X), location (L), medical & health (M) und macro-level Indikatoren (Y). Die letzten vier Zeichen des Variablennamens zeigen das Erhebungsjahr, aus dem die Variable stammt. Diese parallel aufgebaute Struktur erlaubt es den Benutzern, für alle Längsschnittstudien (Panels) die gleichen Programme zu verwenden, idealerweise sogar in einem Durchlauf.

Darüber hinaus erlaubt der spezielle Aufbau des CNEF erfahreneren Forschern schnell und einfach die Algorithmen zu modifizieren, die der Generierung von Variablen zugrunde liegen. Auch weitere Daten können so zur Ergänzung der bestehenden Variablen hinzugefügt werden. Das CNEF-Handbuch beschreibt für jede Variable die für ihre Erstellung verwendeten Algorithmen unter Verwendung der Bezeichnung in den nationalen Originaldaten. Damit Nutzer die vorhandenen CNEF-Daten mit jenen der Originalbefragungen ergänzen können, werden auch die relevanten Identifikatoren für alle Beobachtungseinheiten auf Individual- und Haushaltsebene aus den ursprünglichen Befragungen zur Verfügung gestellt. Dies ermöglicht auch Robustheitsprüfungen der interessierenden Untersuchungsergebnisse bei Modifikation der Definition von Variablen bzw. der für die CNEF-Variablen verwendeten Algorithmen. Schließlich erlaubt es dem Nutzer, gegebenenfalls eine eigene Herangehensweise zur Verbesserung der existierenden Variablen zu entwickeln.

Ergänzend zum oben erwähnten Algorithmus zur Konstruktion von Variablen weist das CNEF jeder Variablen einen sogenannten *Reliability*-Wert zu, der auf die Kompatibilität zwischen den Befragungen und damit auf die Qualität der Vergleichbarkeit hinweist. Lautet der Code „1“ wird von hoher Kompatibilität der Variablen ausgegangen, wohingegen eine „4“ anzeigt, dass die Befragungen über keine vergleichbaren Variablen verfügen. Die Vergabe dieser Codes basiert auf den Erfahrungen der CNEF-Datenproduzenten, direkten Vergleichen der verwendeten Befragungsmethoden und der Kenntnis institutioneller Unterschiede zwischen einzelnen Ländern.

Eine wesentliche Innovation des CNEF ist die Konstruktion von (Einkommens-)Variablen, die weit über das Datenmaterial der ursprünglichen Befragungen hinausgeht (siehe dazu auch den Beitrag von Goebel et al. in diesem Heft). Von besonderer Bedeutung sind

dabei Haushaltsjahreseinkommen vor und nach (sozial-)staatlicher Umverteilung als Ergebnis der Aggregation einer Reihe von Einkommenskomponenten sowie der Schätzung der Abgabenlast der Haushalte aufgrund von direkten Einkommenssteuern und Sozialversicherungsprämien. Viele dieser Werte könnten ohne erheblichen Aufwand seitens des einzelnen Nutzers nicht errechnet werden, wobei die Simulation der Steuerlast eine besonders große Herausforderung darstellt. Dieser innovative Beitrag des CNEF ermöglicht internationale Vergleiche der verfügbaren Einkommen und der staatlichen Umverteilungseffekte, die in dieser Form mit den Daten der Originalbefragungen bestenfalls nur eingeschränkt möglich sind.

Der nötige Aufwand für die Berechnung der Nettoeinkommen variiert dabei von Studie zu Studie. Im Fall von SLID, das auch als nationale Einkommensstichprobe von Statistics Canada dient, sowie dem SHP³ wird die Steuerlast direkt von den Befragten erfragt. In den anderen Studien müssen diese Werte geschätzt werden. Die Steuer- und Abgabensimulationsprogramme für die BHPS-, SOEP- und HILDA-Befragungen wurden von Forschern an den jeweiligen Instituten entwickelt. Das Programm zur Schätzung der Steuerlast in der BHPS (Levy et al. 2006)⁴ wurde von Stephen Jenkins und seinen Koautoren an der Universität von Essex geschrieben und wird auch weiterhin von ihm betreut. Das entsprechende Programm des SOEP wurde ursprünglich von Johannes Schwarze (heute Universität Bamberg) entwickelt und wird von Markus Grabka am DIW Berlin weiter betreut (Schwarze 1995). Der HILDA Survey bedient sich eines von Bruce Headey an der Universität von Melbourne entwickelten Programms. Im Fall des PSID wurden die Zahlen zur Steuerlast intern geschätzt und der Studie dann bei Veröffentlichung beigelegt. Seit 1993 wird die Steuerlast jedoch mithilfe des Programms TAXSIM (siehe Feenberg und Couts 1993)⁵ des National Bureau of Economic Research von Dean R. Lillard an der Cornell Universität geschätzt.

Der Aufwand zur Berechnung des jährlichen Nettoeinkommens ist im Falle des SOEP besonders groß, da in Deutschland der lebensweltlich relevante Einkommensbegriff eher Monatsbezug hat. Insofern wird im SOEP die relevante Information zum Einkommensbezug im Vorjahr erhoben, indem – getrennt nach mehr als 20 Komponenten – die Zahl der Monate mit Bezug einer gegebenen Komponente zusammen mit dem monatlichen Durchschnittsbetrag erfragt wird. International vergleichbare Angaben zum Jahreseinkommen ergeben sich durch Multiplikation dieser Faktoren und Aggregation der diversen Komponenten. Auf diesem Brutto-Jahreseinkommen aufbauend generiert das Steuer-Simulationsprogramm unter Berücksichtigung aller relevanten Parameter (wie z. B. Splitting, Freibeträge, Steuerprogression) eine geschätzte Steuer- und Sozialversicherungsabgabenlast für alle SOEP-Haushalte.

Ziel für alle CNEF-Datenbestände ist die Konstruktion eines für international vergleichende Wohlfahrtsanalysen relevanten (Vor-)Jahreseinkommens nach Umverteilung durch Steuern und Transfers, vereinfachend als „*household post-government income*“ bezeich-

3 Im SHP müssen allerdings Sozialversicherungsabgaben auf das Arbeitseinkommen von den SHP-Datenproduzenten geschätzt werden.

4 Datensätze zum Nettoeinkommen können direkt vom UK Data Archive abgerufen werden (www.data-archive.ac.uk/findingData/snDescription.asp?sn=3909/mrdoc/pdf/3909userguide.pdf).

5 Butrica und Burkhauser (1997) bieten eine detaillierte Besprechung der Steuer-Kalkulations-Algorithmen des NBER und der PSID. Darüber hinaus vergleichen sie die PSID-Schätzungen, die mithilfe von TAXSIM entstanden sind, mit jenen aus den Jahren 1980–1992.

net.⁶ Dieses ergibt sich aus der Summe von Erwerbseinkommen aus selbständiger und abhängiger Beschäftigung (inklusive eventueller Einmalzahlungen wie Weihnachts- und Urlaubsgeld, Gratifikationen, Boni etc.), Kapitalerträgen, Einnahmen aus Vermietung und Verpachtung, privaten Transfers, öffentlichen Transfers sowie allen anderen monetären Zuflüssen aller Mitglieder eines Haushaltes, abzüglich der Einkommensteuer und Sozialversicherungsabgaben. Darüber hinaus steht in einer separaten Variablen auch eine der wichtigsten nichtmonetären Einkommenskomponenten privater Haushalte, der fiktive Mietwert selbstgenutzten Wohneigentums („*imputed rent*“), zur Verfügung. Alle Variablen auf der Haushaltsebene werden jeweils den einzelnen Mitgliedern eines Haushaltes zugeordnet.

Anhang 1 listet für alle aktuell im CNEF enthaltenen Variablen den jeweiligen Namen, eine Kurzbeschreibung und die Verfügbarkeit in den nationalen Datenbasen. Darüber hinaus enthalten die Code-Bücher des CNEF neben dem oben genannten variablen-spezifischen *reliability code* auch Makroinformationen für jedes Land, wie den jährlichen Lebenshaltungskostenindex. Da diese Daten für alle Befragten eines Landes zu einem gegebenen Zeitpunkt konstant sind, werden sie nur im Code-Buch aufgeführt. Anhang 2 zeigt den Stichprobenumfang in den nationalen CNEF-Datensätzen.

4 Vergleichende Darstellung der nationalen Panelstudien im CNEF

Alle sechs im CNEF enthaltenen Längsschnittstudien erfassen unter anderem Informationen zu Haushaltsstruktur, Einkommen, Erwerbstätigkeit und demografischen Merkmalen. Die Vergleichbarkeit der Mikrodaten wird dabei sowohl von Unterschieden zwischen den verschiedenen Befragungen – zum Beispiel bezüglich der Art und Weise der Erhebung – als auch von intertemporalen Variationen innerhalb einer gegebenen Befragung beeinflusst.

Um einen Eindruck der generellen Vergleichbarkeit der Mikrodaten der sechs nationalen Panels zu vermitteln, zeigt Übersicht 1 eine synoptische Darstellung zentraler Merkmale rund um Erhebungsmethode, Design und Umfang der diversen Stichproben, Weiterverfolgsregeln, Teilnahmequoten, Feldarbeit etc. Mit Ausnahme des SLID verfolgen alle CNEF-Befragungen die Mitglieder der Haushalte der Startwelle der jeweiligen Befragung inklusive deren Nachkommen.⁷ Die Panels unterscheiden sich jedoch bezüglich der Definition der zu befragenden Mitglieder eines Haushaltes. So werden zwar in BHPS, SOEP, HILDA und dem SHP alle Erwachsenen eines Haushaltes befragt, jedoch variiert das relevante Befragungsalter zwischen 14 und 17 Jahren. Alle Befragungen erheben Informationen zu Personen, die in einen bestehenden Stichproben-Haushalt einziehen. BHPS, SOEP, HILDA und SHP erlangen diese Informationen auf direktem Wege, da sie jährlich

⁶ Im Allgemeinen richtet sich die Definition der Einkommensvariablen nach den Empfehlungen der „Canberra Group on Household Income Measurement“ (Canberra Group 2001). Aufgrund von „item-non-response“ fehlende Einkommenswerte werden in den CNEF-Daten unter Zuhilfenahme verfügbarer Längsschnittdaten imputiert. Aufbauend auf einer vergleichenden Analyse von CNEF-Daten für Deutschland, Australien und Großbritannien zum Einfluss von Imputation auf Ergebnisse zur Einkommensverteilung und -mobilität argumentieren Frick und Grabka (2007) für eine stärkere Harmonisierung der Imputationstechniken in international vergleichend aufgebauten Datenbasen.

⁷ In der SLID-Studie werden nur die ursprünglichen Mitglieder eines Haushaltes über einen Zeitraum von maximal sechs Jahren weiterverfolgt. Nachkommen dieser Haushalte werden nicht berücksichtigt, sobald sie den gemeinsamen Haushalt verlassen. Insofern sind diese nur im Querschnitt von Bedeutung, im Längsschnitt ist ihr Gewicht gleich null. Gemäß des für diese Studie geltenden Rotationsprinzips beginnt alle drei Jahre ein neues Panel, das somit zu einem gegebenen Zeitpunkt (=Welle) jeweils etwa die halbe Stichprobe repräsentiert.

Übersicht 1

Synopse zentraler Merkmale der Panelstudien im CNEF

Merkmalskategorie	PSID	SOEP	BHPS	SLID	HILDA Survey	SHP
Institut	Institute for Social Research, University of Michigan	SOEP; Deutsches Institut für Wirtschaftsforschung (DIW Berlin)	Institute for Social and Economic Research, University of Essex	Statistics Canada	Melbourne Institute of Applied Economic and Social Research, University of Melbourne	Schweizer Haushalt-Panel, Universität Neuchâtel (seit 2008; Forschungszentrum Sozialwissenschaften (FORS), Universität Lausanne)
Finanzierung	National Science Foundation, National Institute of Health, sowie einige andere Organisationen ¹	1984 bis 2002: Deutsche Forschungsgemeinschaft (DFG); und seit 2000: Bundesministerium für Bildung und Forschung (BMBF). Seit 2003: Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL) ²	UK Economic and Social Research Council	Statistics Canada	Australian Government Department of Families, Community Services and Indigenous Affairs. Reserve Bank of Australia (für Welle 2 im Jahr 2002)	Schweizerischer Nationalfonds
Studiendesign	Unbegrenzt laufende Längsschnittstudie	Unbegrenzt laufende Längsschnittstudie	Unbegrenzt laufende Längsschnittstudie	Überlappende Längsschnittstudien von jeweils sechsjähriger Dauer	Unbegrenzt laufende Längsschnittstudie	Unbegrenzt laufende Längsschnittstudie
Jahr des ersten Interviews	1968	1984	1991	1993	2001	1999
Referenz-Population / Erhebungseinheit	Haushaltsvorstände, deren Haushalt seit mindestens zwei Jahren ohne Unterbrechung in den USA lebt.	Alle privaten Haushalte. Alle Mitglieder ab dem Kalenderjahr, in dem sie 17 Jahre alt werden.	Alle privaten Haushalte. Alle Mitglieder im Alter von 16 und mehr Jahren.	Private Haushalte der zehn Provinzen, mit Ausnahme der Indianerreservate. Alle Mitglieder im Alter von 16 und mehr Jahren. Proxy-Interviews zulässig.	Alle privaten Haushalte, mit Ausnahme derer in sehr abgeschiedener Lage. Alle Mitglieder im Alter von 15 und mehr Jahren.	Alle privaten Haushalte. Alle Mitglieder im Alter von 14 und mehr Jahren.
Erhebungsmethode	Wellen 1–5 (1968–1972) PAPI. Seit Welle 6 (1973) hauptsächlich Telefon. Seit Welle 26 (1993) CATI.	Wellen 1–14 (1984–1997) PAPI. Seit Welle 2 (1985) Methoden-Mix (face-to-face und Selbstausfüller). Seit Welle 15 (1998) Übergang zu CAPI.	Wellen 1–9 (1991–1999) PAPI sowie ein kurzer selbst auszufüllender Fragebogen. Seit Welle 10 (2000) CAPI. Seit Welle 3 (1993) kurzes Telefon-Interview als letzte Möglichkeit.	Seit Welle 1 (1993) CATI.	Seit Welle 1 (2001) PAPI sowie ein kurzer selbst auszufüllender Fragebogen. Kurzes Telefon-Interview als letzte Möglichkeit.	Seit Welle 1 (1999) CATI.

Merkmal	PSID	SOEP	BHPS	SLID	HILDA Survey	SHP
Weiterverfolgungsregeln	Mitglieder der Stichprobe in Welle 1 und deren Nachkommen und adoptierten Kinder. Informationen über Personen, die mit einem ursprünglichen Mitglied der Stichprobe oder mit deren Nachkommen oder adoptierten Kindern zusammenleben.	Mitglieder der Stichprobe in Welle 1 und deren Nachkommen und adoptierten Kinder. Seit Welle 5 (1988) werden Personen, die mit einem ursprünglichen Stichprobenmitglied zusammengewohnt haben, im Falle eines Umzugs innerhalb Deutschlands ebenfalls weiterverfolgt.	Mitglieder der Stichprobe in Welle 1 und deren Nachkommen und adoptierten Kinder. Personen, die mit einem ursprünglichen Mitglied zusammenwohnen werden zu Mitgliedern der aktuellen Welle. Personen, die ein Kind mit einem ursprünglichen Mitglied haben, werden permanente Mitglieder.	Mitglieder der Stichprobe in Welle 1.	Mitglieder der Stichprobe in Welle 1 und deren Nachkommen und adoptierten Kinder. Personen, die mit einem ursprünglichen Mitglied zusammenwohnen werden zu Mitgliedern der aktuellen Welle. Personen, die ein Kind mit einem ursprünglichen Mitglied haben, werden permanente Mitglieder.	Mitglieder der Stichprobe in Welle 1 und deren Nachkommen und adoptierten Kinder. Personen, die mit einem ursprünglichen Mitglied zusammenwohnen werden zu Mitgliedern der aktuellen Welle. Personen, die ein Kind mit einem ursprünglichen Mitglied haben, werden permanente Mitglieder.
Proxy-Interviews (Erwachsene)	Ja – 100% der Nichthaustaltsvorstände. 1976 und 1985 wurden auch Partner befragt.	Nein – keine	Ja – 2 bis 4 %	Ja – circa 30%	Nein – keine	Ja – 2 bis 3 %
Zahl der HH-Interviews in Startwelle	4 802 Familien	5 921 Haushalte	5 538 Haushalte	15 006 Haushalte	7 682 Haushalte	5 074 Haushalte
Zahl der HH-Interviews in der jüngsten Welle	8002 Haushalte (Welle 34, 2005)	11689 Haushalte (Welle 24, 2007)	8 709 Haushalte (Welle 15, 2005)	38 776 Haushalte (Welle 2 / 4. Panel, 2003)	7 139 Haushalte (Welle 6, 2006)	4 256 Haushalte (Welle 7, 2005)
Oversampling (Überrepräsentation) / Zusatzstichprobe	Welle 1 (1968) – Oversampling einkommensschwacher Haushalte (n=1 872). Ca 2/3 dieser Stichprobe wurden ab 1997 nicht weiterverfolgt. Welle 23 (1990) – „Latino“-Sample (nur bis 1 995 erhoben). Welle 30 (1997) – Zuwanderer top-up Sample	Welle 1 (1984) – Oversampling von Migranten-Haushalten (n=1 393). Welle 7 (1990) – Repräsentative Stichprobe Ostdeutschland (n=2 179 Haushalte) Welle 12 (1995) – Oversampling „Zuwanderer seit 1984“ Wellen 15 (1998), 17 (2000) und 23 (2006) – Repräsentative Ergänzungsstichproben Welle 19 (2002) – Zusatzstichprobe „Einkommensstarke Haushalte“	Welle 7 (1997) – Niedrigeinkommens Stichprobe für ECHP – ab Welle 12 (2002) nicht weiterverfolgt. Welle 9 (1999) – Stichproben für Schottland und Wales Welle 11 (2001) – Stichprobe für Nord-Irland	Stichprobe basiert auf dem Labour Force Survey, wobei kleinere Regionen überrepräsentiert werden	Keine.	Welle 6 (2004) – Repräsentative Ergänzungs- und Aufschichtsstichprobe

Merkmal	PSID	SOEP	BHPS	SLID	HILDA Survey	SHP
Welle 1 Teilnahme-Quote in % aller Haushalte	76%	Jeweils vollständig befragte Haushalte: 1984 Sample A: = 61 % 1984 Sample B: = 68 % 1990 Sample C: = 70 % Jeweils inkl. teilweise realisierte Haushalte 1998 Sample E = 54 % 2000 Sample F = 52 % 2006 Sample H = 41 %	Vollständig realisierte Haushalte = 69% (inkl. Proxy Interviews) Inkl. teilweise realisierte Haushalte = 74 % 1999 Sample Schottland/Wales, inkl. teilweise realisierte Haushalte = 63 % 2001 Sample Nord-Irland, inkl. teilweise realisierte Haushalte = 69%	93 %	Vollständig realisierte Haushalte = 59 % Inkl. teilweise realisierte Haushalte = 66 %	Inkl. teilweise realisierte Haushalte = 49%
Teilnahme-Quote ³ : Welle 5 Welle 10 Welle 15 Welle 20	81 % 70% 61 % 52%	69% (71%) ⁴ 53% (55%) 41% (44%) 31% (35%)	72% ⁵ 62% — —	Teilnahmequote in Welle 5: 82% (Panel 1) 79% (Panel 2) 76% (Panel 3)	74% — — —	56% — — —
Feldarbeit	Datenerfassung extern. Panel-Management und Datenbereinigung intern.	Datenerfassung, Teile des Managements sowie der Datenbereinigung extern. Weitergehende Datenprüfung und Generierung intern.	Datenerfassung extern. Management der Studie und Datenbereinigung intern.	Alles intern.	Datenerfassung, Management und Datenverarbeitung extern.	Datenerfassung extern. Management der Studie und Datenbereinigung intern.
Datenweitergabe	Frei zugänglich (online verfügbar).	DVD-Zugang nur für wissenschaftliche Forschungszwecke. Zusätzlich Fernzugang für zweckgebundene Forschung möglich.	Im UK Data-Archive hinterlegt	Derzeit nur über Fernzugang oder Vor-Ort Zugang bei Statistics Canada.	CD-Rom. Zugang nur für wissenschaftliche Forschungszwecke im Rahmen eines definierten Forschungsprojektes.	CD-Rom. Zugang nur für wissenschaftliche Forschungszwecke im Rahmen eines definierten Forschungsprojektes.

1 Die ursprüngliche Finanzierung der PSID erfolgte durch das Office of Economic Opportunity of the United States Department of Commerce. Außerdem haben zur Finanzierung der PSID beigetragen: National Institute on Aging, National Institute of Child Health and Human Development, Offices of the Assistant Secretary for Planning and Evaluation of the United States Department of Health and Human Services, Economic Research Service of the United States Department of Agriculture, United States Department of Housing and Urban Development, United States Department of Labor, Center on Philanthropy at the Indiana University-Purdue University. **2** Die Deutsche Forschungsgemeinschaft (DFG) und die Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL) werden finanziert durch die Bundesregierung und die Länder-Regierungen über die Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (bis 2007) bzw. die Gemeinsame Wissenschaftskonferenz (GWK) von Bund und Ländern (seit 2008). Das DIW Berlin (ebenfalls über die WGL finanziert) trägt einen kleinen Teil der Personalkosten und die laufenden Raum- und Verwaltungskosten. **3** Mit Ausnahme der Angaben für die PSID zeigen diese Teilnahme-Quoten den Anteil der Befragten aus Welle 1, die auch in späteren Wellen erfolgreich befragt wurden. Die Zahlen für die PSID beziehen sich auf die Anzahl der Haushaltsmitglieder aus Welle 1, die auch in späteren Wellen noch im Sample enthalten sind (vgl. Fitzgerald et al., 1998 Tabelle 1). Dies hat zur Folge, dass die Angaben für die PSID in diesem Zusammenhang nur eingeschränkt mit jenen der anderen Studien vergleichbar sind. **4** Die Angaben in Klammern gelten nur für die westdeutsche Stichprobe A. **5** Diese Angaben gelten nur für Befragte mit vollständig absolviertem Interview.

alle erwachsenen Mitglieder des Haushaltes befragen. In der PSID wird immer nur ein Mitglied eines Haushaltes, im Sinne eines Vorstandes oder einer Bezugsperson, befragt, während SLID in großem Maße Proxy-Interviews durchführt. Darüber hinaus unterscheidet sich die SLID-Stichprobe als rotierendes Panel grundsätzlich von den anderen Befragungen, wobei sich die Untersuchungspopulation eines gegebenen Jahres jeweils aus zwei über sechs Jahre laufenden Längsschnittstudien ergibt, die sich über drei Jahre hinweg überschneiden.

Fünf der vorliegenden Befragungen haben im Laufe der Zeit ihre Methoden zur Datenerfassung verändert. Die ursprünglichen Umfragen basierten auf interviewergeführten, persönlichen Interviews, in deren Verlauf der Interviewer die Angaben der Befragten in einen Fragebogen aufnahm („paper and pencil interviewing“ (PAPI)). In den 1990er Jahren stiegen die meisten Befragungen auf computergestützte Methoden um. Die wichtigste Unterscheidung in der Art und Weise der Durchführung computergestützter Interviews bezieht sich darauf, ob diese persönlich („face-to-face“) oder per Telefon durchgeführt werden. BHPS, SOEP und der HILDA Survey werden überwiegend „face-to-face“ durchgeführt. BHPS und SOEP bedienen sich dabei jedoch immer häufiger eines Laptops („computer assisted personal interviewing“ (CAPI)). In SOEP und HILDA findet sich auch eine parallele Anwendung verschiedener Methoden, wobei im SOEP Selbstausfüller („self-completion“) eine immer größere Rolle spielen. HILDA setzt verstärkt Telefon-Interviews ein, da – bedingt durch Umzüge seit der Stichprobenziehung – die regionale Clusterung der zu befragenden Haushaltsadressen abnimmt, was zu entsprechend höheren (Anreise-)Kosten der Interviewer führt. Fast 7 % der von HILDA in der 6. Welle durchgeführten Interviews wurden telefonisch abgewickelt. Die PSID ging 1973 von interviewergestützten („face-to-face“), schriftlichen („paper and pencil interviews“) zu telefonischen Interviews über und wechselte schließlich 1993 zu computergestützten Telefon-Interviews („computer-assisted telephone interviewing“ (CATI)). Sowohl SLID (seit 1992) als auch SHP (seit 1999) verwenden von Anfang an ein CATI-System.

Der Befragungszeitraum, während dem die Befragungen innerhalb eines jeweiligen Jahres durchgeführt werden, variiert zwischen den Studien.⁸ Je nach Art und Zielsetzung einer Analyse sind somit saisonale Effekte als Erklärung für internationale Unterschiede nicht auszuschließen. Die Datenerhebung in SOEP und PSID erfolgt primär in den ersten vier Monaten des Jahres, während sich BHPS auf die Herbstmonate konzentriert. Im Fall von HILDA findet das Gros der Befragungen zwischen September und Dezember (also im australischen „Frühsommer“) statt, für SHP sind es die Wintermonate September bis Februar. Zum Teil sind diese Unterschiede zurückzuführen auf nationale Unterschiede in der Definition des fiskalischen bzw. des Haushaltsjahres.

Unterschiede zwischen den sechs Befragungen gibt es im Zusammenhang mit der Teilnahmewahrscheinlichkeit sowohl bezüglich der Beteiligung in Welle 1 als auch – in etwas geringerem Maße – beim Ausfallverhalten im weiteren Verlauf der Längsschnittstudie („Panel Attrition“). Je nach Definition beträgt die durchschnittliche Teilnahme-Quote der sechs Befragungen in der Startwelle circa 70%. Der Anteil vollständig realisierter Haushalte (also Haushalte, in denen alle Befragungspersonen auch erfolgreich befragt werden konnten) variiert von rund 50% im Fall des SHP⁹, über 59% im HILDA Survey und bis

⁸ In ihren Anfängen haben alle sechs Befragungen die Daten jährlich erfasst. PSID wechselte jedoch 1997 von einem jährlichen zu einem zweijährlichen Rhythmus.

⁹ Zu beachten ist hierbei, dass aufgrund der im SHP angewandten CATI-Methode jeder nicht erfolgreich kon-

zu 76 % in der PSID. Die entsprechenden Werte für BHPS und SOEP liegen bei respektive 69 % und 65 %.¹⁰ Da bei der PSID-Ehebungsmethode nur die Beteiligung eines einzigen Haushaltsmitgliedes nötig ist, schneiden BHPS und SOEP auch im direkten Vergleich mit der US-Studie recht gut ab. In jüngeren Stichproben ist ein allgemein rückläufiger Trend der Teilnahmewahrscheinlichkeit in der Startwelle erkennbar. Dies gilt sowohl im internationalen Vergleich als auch, wie am Beispiel der Auffrischungsstichproben des SOEP erkennbar, innerhalb der einzelnen Länder.

Der Anteil der Panelausfälle stabilisiert sich in den meisten Längsschnittstudien nach einigen Wellen auf recht niedrigem Niveau (circa 4 % oder weniger). Dies trifft auch auf die meisten der im CNEF vertretenen Panels zu. Der Anteil der Personen aus der Startwelle, die bis zur 5. Welle weiterhin teilnehmen, bewegt sich bei SOEP, BHPS und HILDA zwischen 71 % und 74 %. Im SHP ist die Teilnahmewahrscheinlichkeit bis Welle 5 mit lediglich 56 % deutlich niedriger, während die entsprechenden Quoten bei SLID und PSID mit mehr als 80 % überdurchschnittlich hoch ausfallen. Diese höheren Teilnahme-Quoten liegen zumindest teilweise daran, dass die PSID nur jeweils ein Haushaltsmitglied befragt bzw. die SLID-Erhebung den Mitgliedern eines Haushaltes erlaubt, stellvertretend füreinander zu antworten (circa 30 % der Antworten erfolgen im Rahmen solcher Proxy-Interviews). Die Zahl der Befragten mit ununterbrochener Teilnahme an einer Längsschnittstudie („balanced panel“) schrumpft grundsätzlich mit jedem zusätzlichen Beobachtungsjahr aufgrund demografischer Ereignisse wie Tod und Auswanderung sowie – oft selektivem – Ausfallverhalten (*attrition*). Diese negativen Entwicklungen können zumindest im Hinblick auf die Stichprobengröße bei Querschnittuntersuchungen durch die Geburt neuer Haushaltsmitglieder und Zuzüge in bereits teilnehmende Stichprobenhaushalte ausgeglichen werden.

Die Befragungen unterscheiden sich in Bezug auf Erweiterungen der Stichprobe und die Anwendung von „oversampling“, also bewusst überrepräsentierenden Teilstichproben von Untersuchungseinheiten mit eher selten vorkommenden Charakteristika. Nachdem Kritik an der mangelnden Repräsentation von Migranten in den Daten der PSID laut geworden war, wurde 1990 eine Stichprobe von Lateinamerikanern („Latino-Sample“) hinzugenommen, die jedoch später wieder aufgegeben wurde, sowie eine nicht herkunftslandspezifische Zuwandererstichprobe, die bis heute besteht. Besonderes Merkmal des SOEP ist eine kontinuierliche Erweiterung der zu befragenden Population durch eine Reihe von repräsentativen Auffrischungs- und Ergänzungsstichproben (diese fanden bisher in den Jahren 1998, 2000 und 2006 statt). Darüber erfolgte im Juni 1990, nur wenige Monate nach dem Mauerfall und noch vor Inkrafttreten der Währungs-, Wirtschafts- und Sozialunion, die Ausdehnung des SOEP-Erhebungsgebietes auf Ostdeutschland. Auch das „oversampling“ bestimmter Bevölkerungsgruppen mit hoher Relevanz für sozialstrukturelle Analysen ist ein Charakteristikum des SOEP. Bei solchen Teilgruppen handelt es sich insbesondere um Personen in Haushalten mit Migrationshintergrund (in den Wellen 1984 und 1995) und einkommensstarke Haushalte (im Jahr 2002). Auch die BHPS hat im Rahmen der Integration des UK-ECHP-Samples eine Teilstichprobe von Haushalten mit niedrigem Einkommen aufgenommen, diese aber später wieder aufgegeben. Für die Befragungen HILDA und SLID ist das Thema der Wiederauffrischung der Stichprobe von etwas geringerer Bedeutung, da SLID sich auf die Überlappung relativ kurz angelegter Längsschnittstu-

taktierte Haushalt als Ausfall gilt, was qualitätsneutrale Ausfälle einschließt.

10 Die Teilnahmequoten in der ersten Welle der beiden ursprünglichen Teilstichproben des SOEP betragen 61 % bei den „West-Deutschen“ (Sample A) und 68 % im „Ausländer“-Sample B.

dien beschränkt und der HILDA Survey noch verhältnismäßig jung ist. Das SHP ist zwar ebenfalls noch relativ jung, reagierte aber bereits im Jahre 2004 mit einer repräsentativen Auffrischungsstichprobe auf die überdurchschnittlichen Verluste durch Panel Attrition.

Es gibt diverse gute Gründe für repräsentative Stichprobenergänzungen insbesondere in lang laufenden Längsschnittstudien. Solche Erweiterungen dienen nicht nur zur Vergrößerung der Stichprobe bzw. zur Stabilisierung der Fallzahlen im Querschnitt, sondern sind auch Basis für die empirische Untersuchung von Befragungs- oder Paneleffekten in länger laufenden Stichproben (vgl. z. B. Frick et al. 2006). Darüber hinaus können solche Auffrischungsstichproben demografische Veränderungen in der Grundgesamtheit (aufgrund von Immigration nach dem Zeitpunkt der Stichprobenziehung) nachbilden und somit entsprechende Verzerrungen der bereits vorliegenden Stichproben ausgleichen.

Die Ziehung solcher Ergänzungsstichproben ist somit neben dem Saldo der „natürlichen“ Stichprobenentwicklung (Verluste aufgrund demografisch bedingter Ausfälle und Verweigerungen bzw. Gewinne aufgrund von Abspaltungen aus bestehenden Stichproben-Haushalten) die entscheidende Einflussgröße für die Abweichung der befragungsspezifischen Stichprobengröße in der aktuellsten Welle im Vergleich zur jeweiligen Startwelle. Dies kann am Beispiel des SOEP anschaulich gemacht werden. Im Rahmen der ersten Welle im Jahre 1984 wurden knapp 6 000 Haushaltsinterviews (Samples A und B) durchgeführt. Bedingt durch die Ergänzungen um sechs weitere Sub-Stichproben (Samples C bis H) sowie das Entstehen „neuer“ Haushalte, die durch räumliche Mobilität bzw. Abspaltungen aus den ursprünglichen Haushalten hervorgingen, konnten im Jahr 2006 rund 12 500 Haushalte interviewt werden.

Schließlich unterscheiden sich die Studien auch recht deutlich bezüglich der Organisation und Durchführung von Feldarbeit und Datenproduktion. Da SLID die zentrale Einkommensstichprobe des nationalen statistischen Amtes (Statistics Canada) ist, wird der gesamte Prozess der Datenerfassung, -prüfung und -bereinigung intern betreut. Tendenziell ähnlich ist es bei der PSID, wo wissenschaftliche Leitung und Datenbereinigung in ein und demselben wissenschaftlichen Institut vereint sind. Im Falle der anderen Studien erfolgt die Datenerhebung extern durch private Firmen. Codierung und Aufbereitung der Daten für die weitere wissenschaftliche Analyse erfolgt in den meisten der CNEF-Studien *inhouse*. Zwar sind beim SOEP weite Teile der Aufbereitung ebenfalls an das Erhebungsinstitut Infratest Sozialforschung ausgegliedert, jedoch die Generierung nutzerfreundlicher Variablen im Quer- und Längsschnitt, die Imputation fehlender Werte sowie Hochrechnung und Gewichtung finden fast ausschließlich beim SOEP selbst statt.

Die Auswahl der hier dargestellten Unterschiede und Gemeinsamkeiten dieser sechs Studien ist offensichtlich fokussiert auf die direkt für die CNEF relevanten Aspekte. Weitere Details zu den diversen Studien finden sich in Frick et al. (2007) sowie auf den entsprechenden Web-Seiten:¹¹

- The Panel Study of Income Dynamics, PSID (www.psidonline.isr.umich.edu/)
- Das Sozio-oekonomische Panel, SOEP (www.diw.de/gsoep)
- The British Household Panel Survey, BHPS (www.iser.essex.ac.uk/ulsc/bhps/doc/)

11 Im Literaturverzeichnis finden sich Hinweise für vertiefende Darstellungen der einzelnen Befragungen.

- The Survey of Labour and Income Dynamics, SLID (www.statcan.ca/start.html)
- The Household, Income and Labour Dynamics in Australia Survey, HILDA (<http://melbourneinstitute.com/hilda>)
- Schweizerisches Haushalt-Panel, SHP (www.swisspanel.ch)

Zukünftige Entwicklungen dieser und neuer nationaler Befragungen werden in erster Linie die Informations- und Analysebedürfnisse der nationalen politischen Entscheidungsträger und Forschungsstrategien reflektieren. Somit bleibt die Notwendigkeit einer kontinuierlichen Harmonisierung auch der zukünftig noch zu erhebenden nationalen Daten für dieses international vergleichende Projekt erhalten.

5 Ausblick und Zugang zu den CNEF-Daten

Die Daten des CNEF erlauben Forschern mit Interesse an sozio-ökonomischen Fragestellungen, international vergleichende Analysen im Quer- und Längsschnitt für Australien, Kanada, Deutschland, Großbritannien, die Schweiz und die USA durchzuführen. Im Gegensatz zu anderen Querschnittsdatenbanken enthält das CNEF nicht nur detaillierte Angaben darüber, wie die vorliegenden Variablen generiert wurden, sondern unterstützt Wissenschaftler auch bei entsprechenden Modifikationen und Erweiterungen, da die CNEF-Daten mit den Original-Befragungsdaten verknüpfbar sind. Die bisher auf Basis des CNEF vorgelegten Analysen bieten insofern auch Anwendungsbeispiele für zukünftige Weiterentwicklungen im Sinne einer forschungsgetriebenen Verbesserung des Datenangebotes. Die Prozesse zur Harmonisierung bereits bestehender Variablen werden dabei kontinuierlich überprüft und gegebenenfalls optimiert. Derzeit enthält das CNEF zwar nur einen Bruchteil der Variablen der nationalen Panels. Laufende Forschungsarbeiten der internationalen Nutzergemeinschaft, die kontinuierlich neue Forschungsbereiche aufgreifen, führen aber auch zur Generierung von mit großer Sorgfalt befragungsübergreifend definierten Daten. Das jüngste Beispiel für diese Art der Erschließung neuer Themenfelder ist der Bereich Gesundheit.¹² Gleichzeitig wird die verbesserte Zusammenarbeit zwischen Datenproduzenten und Analysten, die derzeit gemeinsam an der nachträglichen, also ex-post Harmonisierung der vorliegenden Surveydaten arbeiten, auch die zukünftige ex-ante Harmonisierung neuer Bereiche unterstützen. Dies wird wiederum zur Verbesserung der internationalen Vergleichbarkeit der Mikro-Daten beitragen und so auch die Qualität der Forschungsergebnisse steigern.

Der Zugang zu den CNEF-Daten ist abhängig von nationalen Datenschutzbestimmungen. Da PSID-Daten grundsätzlich öffentlich zugänglich sind, werden die PSID-CNEF-Datensätze auf der CNEF-Webseite ebenfalls der Öffentlichkeit zugänglich gemacht. Der Zugang zu BHPS-CNEF, SOEP-CNEF, HILDA-CNEF und SHP-CNEF-Daten ist nur mit der Erlaubnis des jeweiligen nationalen Datenproduzenten möglich.¹³ Diese Erlaubnis muss dann wiederum, per E-Mail oder Fax, an das CNEF-Büro an der Cornell Universität geschickt werden. Erst dann wird die CNEF-CD zugestellt. Eine Ausnahme bilden die

¹² Als zukünftige Erweiterungen werden derzeit subjektive Indikatoren wie die „Allgemeine Lebenszufriedenheit“ diskutiert sowie zusätzliche „nichtmonetäre“ Einkommenskomponenten zur Ergänzung der bereits verfügbaren Messung des fiktiven Mietwerts für selbstnutzende Eigentümer. Diese Informationen werden in der Variablen „Imputed Rental Value of owner-occupied housing“ (Variable I11105_xxxx) zur Verfügung gestellt.

¹³ Es handelt sich hierbei jedoch um eine relativ einfache und schnell abgewickelte Formalität. Eine detaillierte Beschreibung der relevanten Schritte findet sich auf der CNEF-Webseite der Cornell Universität.

SLID-Daten, die sich aufgrund der kanadischen Datenschutzbestimmungen nicht auf der CNEF-CD befinden. Registrierte CNEF-Nutzer haben die Möglichkeit, im Rahmen einer Datenfernverarbeitung ihre Analyse-Programme an Statistics Canada zu senden. Diese werden dort nach datenschutzrechtlicher Prüfung ausgeführt und die entsprechenden Ergebnisse an den Forscher zurück gesandt.¹⁴

Die einmalige Anmeldegebühr für die Daten des CNEF beträgt 125 US-Dollar, zahlbar an die Cornell Universität. Für weitere Informationen zum Datenzugang konsultieren Sie bitte die CNEF-Webseite unter: www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/cnef.cfm oder senden Sie eine E-Mail an: cnef@cornell.edu

Literaturverzeichnis

- Burkhauser, Richard V., Amy Crews-Cutt und Dean R. Lillard (1999): How Older People in the United States and Germany Fared in the Growth Years of the 1980s: A Cross-Sectional versus a Longitudinal View. *Journal of Gerontology*. 54B (5), S279–S290.
- Burkhauser, Richard V., Joachim R. Frick und Johannes Schwarze (1997): A Comparison of Alternative Measures of Economic Well-Being for Germany and the United States. *Review of Income and Wealth*, 43 (2), 153–171.
- Burkhauser, Richard V., Philip Giles, Dean R. Lillard und Johannes Schwarze (2005): Until Death Do Us Part: An Analysis of the Economic Well-Being of Widows in Four Countries. *Journal of Gerontology: Social Sciences*. 60B (5), S238–S246.
- Burkhauser, Richard V. und Dean R. Lillard (2005): The Contribution and Potential of Data Harmonization for Cross-National Comparative Research. *Journal of Comparative Policy Research*. 7, 313–330.
- Burkhauser, Richard V. und Dean R. Lillard (2007): The Expanded Cross-National Equivalent File: HILDA joins its International Peers. *The Australian Economic Review*. 40, 208–215.
- Burkhauser, Richard V., Maximilian D. Schmeiser und Mathis Schroeder (2007): The Employment and Economic Well Being of Working-Age Men with Disabilities: Comparing Outcomes in Australia, Germany, and Great Britain with the United States. Paper presented at the HILDA Survey Research Conference 2007, July 19-20 2007, University of Melbourne.
- Butrica, Barbara A. und Richard V. Burkhauser (1997): Estimating Tax Burdens in the PSID Using the TAXSIM Model, Aging Studies Program Project Paper.12.
- Butz, William B. und Barbara B. Torrey (2006): Some frontiers in Social Sciences. *Science*, 312, 1898–1900.
- Canberra Group (2001): *Expert Group on Household Income Statistics: Final Report and Recommendations*. Ottawa.
- Feenberg, Daniel und Elisabeth Coutts (1993): An Introduction to the TAXSIM Model. *The Journal of Policy Analysis and Management*, 12, 189–194.
- Frick, Joachim R., Jan Goebel, Edna Schechtman, Gert G. Wagner und Shlomo Yitzhaki (2006): Using Analysis of Gini (ANoGi) for detecting whether two sub-samples repre-

¹⁴ Im Rahmen kürzlich geänderter Bestimmungen ist der Zugang zu den SLID-Daten nun auch im Rahmen eines persönlichen Forschungsaufenthaltes in einem der Research Data Centres (RDC) in Kanada möglich. Dieses Netzwerk umfasst 14 Forschungsdatenzentren, sechs Zweigstellen und das staatliche Forschungszentrum in Ottawa.

- sent the same universe: The German Socio-Economic Panel Study (SOEP) Experience. *Sociological Methods & Research*, 34 (4), 427–468.
- Frick, Joachim R. und Markus Grabka (2007): *Item Non-Response and Imputation of Annual Labor Income in Panel Surveys from a Cross-National Perspective*. IZA Discussion Paper. 3043. <http://ftp.iza.org/dp3043.pdf>
- Frick, Joachim R., Stephen P. Jenkins, Dean R. Lillard, Oliver Lipps und Mark Wooden (2007): The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies. *Schmoller's Jahrbuch – Journal of Applied Social Science Studies*. 127 (4), 627–654.
- Haisken-DeNew, John und Joachim R. Frick (2005): Desktop Companion to the German Socio-Economic Panel Study (GSOEP). Version 8.0 – Update to Wave 21. DIW Berlin.
- Headey, Bruce (2003): *How Best to Impute Taxes and Measure Public Transfers?* HILDA Discussion Paper Series. 2/03.
- Jenkins, Stephen P. und Chris Schluter (2003): Why are Child Poverty Rates Higher in Britain than in Germany? A Longitudinal Perspective. *Journal of Human Resources*, 38, 441–465.
- Jenkins, Stephen P., Chris Schluter und Gert G. Wagner (2003): The Dynamics of Child Poverty: Britain and Germany Compared. *Journal of Comparative Family Studies*, 34, 337–353.
- Levy, Horacio, Francesca Zantomio, Holly Sutherland und Stephen P. Jenkins (2006): Documentation for Derived Current and Annual Net Household Income Variables, BHPS Waves 1-14. [Retrieved from: www.data-archive.ac.uk/doc/3909/mrdoc/pdf/3909userguide.pdf]
- Lillard, Dean R. und Richard V. Burkhauser (2006): *Evaluation of the Cross-National Comparability of the Survey of Health, Ageing, and Retirement in Europe, the Health and Retirement Study, and the English Longitudinal Study of Ageing*. Report prepared for the National Institutes of Aging, Washington, D. C.
- Lillard, Dean R. und Gert G. Wagner (2006): The Value Added of Biomarkers in Household Panel Studies. DIW Data Documentation. 14.
- Schwarze, Johannes (1995): Simulating the Federal Income and Social Security Tax Payments of German Households Using Survey Data. Cross-National Studies in Aging Program Project Paper. 19.
- Smeeding, Timothy M., David K. Jesuit und Paul Alkemade (2002): The LIS / LES Project Databank: Introduction and Overview. *Schmollers Jahrbuch*, 122, 497–717.
- Wagner, Gert G., Joachim R. Frick und Jürgen Schupp (2007): The German Socio-Economic Panel Study (SOEP) – Evolution, Scope and Enhancements. *Schmollers Jahrbuch*, 127, 139–169.
- Watson, Nicole (2006): Options for a Top-up Sample to the HILDA Survey. Paper presented at the ACSPRI Social Science Methodology Conference, University of Sydney, 10–13 December.
- Zimmermann, Erwin und Robin Tillmann (eds.) (2004): *Vivre en Suisse 1999–2000 – Leben in der Schweiz 1999–2000. Une année dans la vie des ménages et familles en Suisse. Ein Jahr im Leben der Schweizer Familien und Haushalte*. Bern, Peter Lang Verlag AG. (Vol. 3. Population, Famille et Société)

Anhang

Tabelle A1

Variablen im Cross-National Equivalent File (CNEF) 1980–2005

Label	Data	Variable Name
Demographie (Demographics):		
Age of Individual	B, G, H, P, S, CH	D11101_XXXX
Sex of Individual	B, G, H, P, S, CH	D11102LL
Marital Status of Individual	B, G, H, P, S, CH	D11104_XXXX
Relationship to Household Head	B, G, H, P, S, CH	D11105_XXXX
Number of Persons in Household	B, G, H, P, S, CH	D11106_XXXX
Number of Children in Household	B, G, H, P, S, CH	D11107_XXXX
Education With Respect to High School	G, H, P, S, CH	D11108_XXXX
Number of Years of Education	G, H, P, S, CH	D11109_XXXX
Race of Individual ¹	B, P, S	D11112LL
Erwerbstätigkeit (Employment):		
Annual Work Hours of Individual	B, G, H, P, S, CH	E11101_XXXX
Impute Annual Work Hours of Individual	B, CH	E111201_XXXX
Employment Status of Individual	B, G, H, P, S, CH	E11102_XXXX
Employment Level of Individual	B, G, H, P, S, CH	E11103_XXXX
Primary Activity of Individual	B, G, P, S, CH	E11104_XXXX
Occupation of Individual	B, G, H, P, S, CH	E11105_XXXX
1 Digit Industry Code of Individual	B, G, H, P, S, CH	E11106_XXXX
2 Digit Industry Code of Individual	B, G, H, P, S, CH	E11107_XXXX
Input zur Berechnung von Äquivalenzskalen (Equivalence Scale inputs):		
Number HH members age 0-14	B, G, H, P, S, CH	H11101_XXXX
Number HH members age 15-18	B, G, H, P, S, CH	H11102_XXXX
Number HH members age 0-1	B, G, H, P, S, CH	H11103_XXXX
Number HH members age 2-4	B, G, H, P, S, CH	H11104_XXXX
Number HH members age 5-7	B, G, H, P, S, CH	H11105_XXXX
Number HH members age 8-10	B, G, H, P, S, CH	H11106_XXXX
Number HH members age 11-12	B, G, H, P, S, CH	H11107_XXXX
Number HH members age 13-15	B, G, H, P, S, CH	H11108_XXXX
Number HH members age 16-18	B, G, H, P, S, CH	H11109_XXXX
Number HH members age 19+ or 16-18 and indep.	B, G, H, P, S, CH	H11110_XXXX
Indicator - Wife/spouse in HH	B, G, H, P, S, CH	H11112_XXXX
(Vor-)Jahreseinkommen (Yearly Income):		
Household Pre-Government Income	B, G, H, P, S, CH	I11101_XXXX
Household Post-Government Income	B, G, H, P, S, CH	I11102_XXXX
Household Labor Income	B, G, H, P, S, CH	I11103_XXXX
Household Asset Income	B, G, H, P, S, CH	I11104_XXXX
Household Imputed Rental Value	B, G, H, P, S, CH	I11105_XXXX
Household Private Transfers	B, G, H, P, S, CH	I11106_XXXX

Label	Data	Variable Name
Household Public Transfers	B, G, H, P, S, CH	I11107_xxxx
Household Social Security Pensions	B, G, P, S, CH	I11108_xxxx
Total Household Taxes	B, G, H, P, S, CH	I11109_xxxx
Individual Labor Earnings	B, G, H, P, S, CH	I11110_xxxx
Household Federal Taxes	G, P	I11111_xxxx
Household Social Security Taxes	B, G, P, CH	I11112_xxxx
Household Post-Government Income (TAXSIM)	P	I11113_xxxx
Total Household Taxes (TAXSIM)	P	I11114_xxxx
Household State Taxes (TAXSIM)	P	I11115_xxxx
Household Federal Taxes (TAXSIM)	P	I11116_xxxx
Household Private Retirement Income	B, G, H, P, S	I11117_xxxx
Household Windfall Income	B, G, H, P, S, CH	I11118_xxxx
Impute Household Pre-Government Income	B, G, H, CH	I11201_xxxx
Impute Household Post-Government Income	B, G, H, CH	I11202_xxxx
Impute Household Labor Income	B, G, H, CH	I11203_xxxx
Impute Household Asset Income	B, G, H, CH	I11204_xxxx
Impute Household Imputed Rental Value	B, G, CH	I11205_xxxx
Impute Household Private Transfers	B, G, H, CH	I11206_xxxx
Impute Household Public Transfers	B, G, CH	I11207_xxxx
Impute Household Social Security Pensions	B, G, CH	I11208_xxxx
Impute Total Household Taxes	G, H, CH	I11209_xxxx
Impute Individual Labor Earnings	B, G, H, CH	I11210_xxxx
Impute Private Retirement Income	B, G, H	I11217_xxxx
Regionalinformation (Location):		
Area of Residence ²	B, G, P, S, CH	L11101_xxxx
Region of Residence	B, G, H, CH	L11102_xxxx
Gesundheit (Medical/health):		
Whether spent night in hospital in last year	B, G, P, CH	M11101_xxxx
Number of nights (days) spent in hospital	B, G, P, CH	M11102_xxxx
Whether had accident in past year that required hospital	B, G, CH	M11103_xxxx
Frequency of sports or exercise	B, G, P, CH	M11104_xxxx
Have had stroke	B, P	M11105_xxxx
Have or had high blood pressure/hypertension	B, P	M11106_xxxx
Have or had diabetes	B, P	M11107_xxxx
Have or had cancer	B, P	M11108_xxxx
Have or had psychiatric problems	B, P	M11109_xxxx
Have or had arthritis	B, P	M11110_xxxx
Have or had angina or heart condition	B, P	M11111_xxxx
Have or had asthma or breathing difficulties	B, P	M11112_xxxx
Have trouble climbing stairs	B, G, P	M11113_xxxx
Have trouble with bath	B, P	M11114_xxxx
Have trouble dressing	B, G, P	M11115_xxxx
Have trouble getting out of bed	B, G, P	M11116_xxxx

Label	Data	Variable Name
Have trouble shopping	G, P	M11117_XXXX
Have trouble walking	B, P	M11118_XXXX
Have trouble doing housework	B, G, P	M11119_XXXX
Have trouble bending, lifting, stooping	B, P	M11120_XXXX
Health limits vigorous physical activities	B, P	M11121_XXXX
Height (in meters)	G, P, CH	M11122_XXXX
Weight (in kilos)	G, P, CH	M11123_XXXX
Disability Status of Individual	B, G, H, P, S	M11124_XXXX
Subjective Satisfaction with Health	B, G, H, S, CH	M11125_XXXX
Self-Rated Health Status	B, G, H, P, CH	M11126_XXXX
Number of Times Visited Dr. in Past Year	G, CH	M11127_XXXX
Hochrechnung und Gewichtung (Weights):		
Cross-sectional Weight - Respondent Individuals	B, G, H, P, S, CH	W11101_XXXX
Household Weight	B, G, H, P, S, CH	W11102_XXXX
Longitudinal Weight - Respondent Individuals	B, G, H, P, S, CH ⁴	W11103_XXXX
Population Factor for W11101_XXXX	B, G, P	W11104_XXXX
Individual Weight - Immigrant Sample	G	W11105_XXXX
Household Weight - Immigrant Sample	G	W11106_XXXX
Cross-sectional Weight - Enumerated Individuals	B, H	W11107_XXXX
Longitudinal Weight - Enumerated Individuals	B, H	W11108_XXXX
Population Factor for W11103_XXXX	B, G, P	W11109_XXXX
Population Factor for W11107_XXXX	B	W11110_XXXX
Population Factor for W11108_XXXX	B	W11111_XXXX
Algorithmen zur Ableitung von Äquivalenzskalen (Equivalence Weight Algorithms):		
Detailed Official U.S. Equivalence Weight		
General Official U.S. Equivalence Weight		
Official German Equivalence Weight		
ELES Equivalence Weight		
OECD Equivalence Weight		
McClements Equivalence Weight		
Other Equivalence Weights		
Identifikatoren (Identifiers):		
Unique Person Number	B, G, H, P, S, CH	X11101LL
Household Identification Number	B, G, H, P, S, CH	X11102_XXXX
Individual in Household at Survey	B, G, H, P, S	X11103_XXXX
Oversample Identifier	B, G, P, S	X11104LL
Person in Household Interviewed	B, G, H, CH	X11105_XXXX
Makro-Indikatoren (Macro-level Variables):³		
Consumer Price Index	B, G, P, S	
Median Pre-government Household Income	B, G, P, S	
Median Post-government Household Income	B, G, P, S	
Purchasing Power Parity for East Germany	G	

Anmerkungen:

(B) BHPS: Erhebungsjahre 1991–2004 / (G) SOEP: Erhebungsjahre 1984–2005 / (H) HILDA: Erhebungsjahre 2001–2004 / (P) PSID: Erhebungsjahre 1980–2003 / (S) SLID: Erhebungsjahre 1992–2005 / (CH) SHP: Erhebungsjahre 1999–2005.

1 Die Variable "Race" wird in BHPS und SLID für alle Stichprobenmitglieder erhoben. In den PSID-Daten trifft dies nur für solche Personen zu, die in mindestens einer Welle Haushaltsvorstand oder Partner waren. **2** „Area of residence“ entspricht in der BHPS dem Local Authority District of Residence, dem Bundesland in den SOEP-Daten, dem US-Bundesstaat in der PSID, dem Kanton im SHP. Diese Information ist in den CNEF-Daten des SLID nicht enthalten, aber im Rahmen eines Forschungsaufenthalts bei Statistics Canada verfügbar. Zugang zu Informationen des Local Authority District of Residence in den BHPS-Daten ist im Rahmen spezieller Arrangements mit der University of Essex möglich. **3** Da diese zeitpunktspezifischen Indikatoren auf Makro-Ebene für alle Individuen oder Haushalte eines Erhebungsjahres gelten, sind diese lediglich in den CNEF-Handbüchern gelistet.

Tabelle A2

Fallzahlen der nationalen Panelstudien im CNEF (Individualdaten)

Jahr	PSID	SOEP	BHPS	SLID	HILDA	SHP
1980	18 989	–	–	–	–	–
1981	18 992	–	–	–	–	–
1982	19 246	–	–	–	–	–
1983	19 491	–	–	–	–	–
1984	19 570	15 237	–	–	–	–
1985	19 787	13 747	–	–	–	–
1986	19 615	13 084	–	–	–	–
1987	19 647	12 853	–	–	–	–
1988	19 687	12 253	–	–	–	–
1989	19 669	11 856	–	–	–	–
1990	19 932	17 462	–	–	–	–
1991	19 962	17 094	13 780	–	–	–
1992	20 334	16 801	13 151	40 155	–	–
1993	21 450	16 510	13 104	42 194	–	–
1994	23 620	16 828	12 851	43 717	–	–
1995	23 182	17 252	12 549	88 230	–	–
1996	23 060	16 869	12 720	91 624	–	–
1997	19 132	16 559	15 042	94 125	–	–
1998	–	18 161	14 835	139 508	–	–
1999	19 669	17 417	21 540	94 772	–	10 437
2000	–	30 439	21 602	96 512	–	9 454
2001	20 538	27 481	26 586	141 598	19 914	8 775
2002	–	29 280	23 435	93 680	18 295	7 648
2003	21 277	27 553	22 559	95 792	17 691	6 944
2004	–	26 690	22 105	139 246	17 209	10 666
2005	22 918	25 544	15 627	91 569	17 469	8 550
Insgesamt (Personen * Wellen)	449 767	416 970	261 486	1 292 722	90 578	62 474

Hinweis: Diese Fallzahlen können von jenen in der jeweiligen nationalen Dokumentation abweichen. So sind z. B. im Format des PSID-CNEF das Niedrigeinkommens- und das Latino-Sample ausgeschlossen und der SOEP-CNEF-Datensatz basiert aus Datenschutzgründen lediglich auf einer 95 % Zufallsstichprobe der Daten der ersten Welle.